Institute for Empirical Research in Economics
University of Zurich

Working Paper Series
ISSN 1424-0459

Working Paper No. 198

Existence and Global Attractivity of Stable Solutions in Neural Networks

Patrick Leoni and Pietro Senesi

August 2004

Existence and Global Attractivity of Stable Solutions in Neural Networks*

Patrick Leoni† Pietro Senesi‡

Abstract

The present paper shows that a sufficient condition for the existence of a stable solution to an autoregressive neural network model is the continuity and boundedness of the activation function of the hidden units in the multi layer perceptron (MLP). In addition, uniqueness of a stable solution is ensured by global lipschitzness and some conditions on the parameters of the system. In this case, the stable value is globally stable and convergence of the learning process occurs at exponential rate.

Keywords: Neural Networks, Stable Value

JEL Classification: C45

*We would like to thank B. Allen, C. Ewerhart and C. Nitzsche for interesting conversations. All remaining errors are ours.

†University of Zurich - Institute for Empirical Research in Economics - Winterthurerstrasse 30, CH-8006 Zurich - Switzerland, pleoni@iew.unizh.ch

‡Università di Roma Tor Vergata - Dip.to di Economia e Istituzioni - Via Columbia, 2 - 00133 Roma - Italy - senesi@uniroma2.it

1 Introduction

Time series analysis and forecasting are important applications of neural network (NN) models. Their applications range from fields such as hydrology (see for instance Poff *et al.* [7]) to finance (see Zirilli [8] among many others). The interest in neural network models relies in their ability to provide a non-linear mechanism for both signal processing and expectations formation (see Beale and Jackson [5], or Kohonen [6] for an introduction). The understanding of long-run behavior of such systems is critical when, for instance, a state of the world is to be learned. In this case, existence of a stationary solution and convergence of the system for a large set of initial inputs play a crucial role.

The characterization of the long-run stochastic properties of the *autoregressive neural network* (AR-NN) process driven by additive noise is done in Trapletti *et al.* [4]. The authors investigate the properties of the omega-limit set of the probability distribution of the AR-NN as a stochastic process. In particular, they make use of Markov chain theory in order to identify conditions for ergodicity of the forecasting process and its stationarity.

The present analysis takes the AR-NN as a model of expectations formation, and investigates the stability properties of its deterministic nonlinear NN component. In particular, the object of this paper is to identify conditions on the NN component for existence, uniqueness and global stability of stationary solutions of the model. To do so, we focus on stationary *values* of the system; i.e., inputs to the system that induce outputs of the AR-NN that become constant after some finite time. The motivation for studying stationary values is that every stable output of the AR-NN converges toward a stationary value.

The results are derived from an interplay between conditions on the parameters of the linear AR part of the model and the nonlinear NN component. Existence of a stationary value is established in Theorem 1 and Proposition 2 under boundedness and continuity of the activation function of the hidden units in the multi layer perceptron (MLP), together with some conditions on the parameters of the AR component. Such conditions are minimal; but they do not guarantee uniqueness. Lack of uniqueness is problematic for learning processes in that there is always the risk for the system output to be attracted toward an incorrect stationary value (i.e., a stationary value different from the state of the world to be estimated or learned). Inaccurate long-run learning may occur even when system parameters are correctly estimated. This stems from the nonlinear NN component driving the system to an incorrect stationary solution.

Uniqueness of a stationary value obtains when both the activation function is globally Lipschitz and the system parameters are such that the system output under a constant input is a contraction operator (Theorem 3). Moreover, in this case, Theorem 4 establishes that, for every possible initial values of the system, the AR-NN converges toward the unique stationary value at exponential rate.

The paper is organized as follows: Section 2 reviews the AR-NN model and introduces the relevant definitions and assumptions. Section 3 investigates existence of a solution to the AR-NN expectational model by applying a Brouwer's fixed point argument along the lines of van den Driessche and Zou [1]. Sufficient conditions for uniqueness of the stationary value are given in Section 4, and global stability and speed of convergence in this later case are studied in Section 5. Finally Section 6 ends the paper with a summary of the results and concluding remarks.

2 The model

Our starting point is the autoregressive process of order $p > 1$, and defined by the nonlinear deterministic difference equations

$$y_t = h\left(y_{t-1}, y_{t-2}, ..., y_{t-p}\right). \tag{1}$$

We take any arbitrary p real numbers $(y_1, ..., y_p)$ as initial conditions.

Following standard literature on neural network modeling, we consider the following representation

$$h\left(y_{t-1}, y_{t-2}, ..., y_{t-p}\right) = \sum_{i=1}^{p} \psi_i y_{t-i} + \sum_{k=1}^{q} \beta_k G\left(\sum_{j=1}^{p} \phi_{k,j} y_{t-j}\right),$$

where the vector of weights $(\psi_1, ..., \psi_p)$ is in $\mathbb{R}^p$,where the expression

$$\left(\phi_1^{'}, ..., \phi_p^{'}\right)' = \begin{bmatrix} \phi_{11} & . & . & . & \phi_{1p} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \phi_{q1} & . & . & . & \phi_{qp} \end{bmatrix}$$

is a $(q \times p)$ matrix of real numbers, where $\beta = \left(\beta_1, ..., \beta_q\right)'$ is a $(q \times 1)$ matrix of real numbers, and where G is the activation function.

Thus, we obtain the equivalent representation

$$\begin{aligned} y_t &= \psi' \left(y_{t-1}, y_{t-2}, ..., y_{t-p}\right) \\ &\quad + \sum_{i=1}^{q} \beta_i G\left(\phi_i' \left(y_{t-1}, y_{t-2}, ...\right)\right) + \varepsilon_t \\ &= \psi_1 y_{t-1} + \psi_2 y_{t-2} + ... \\ &\quad + \sum_{i=1}^{q} \beta_i G\left(\phi_i' \left(y_{t-1}, y_{t-2}, ...\right)\right) + \varepsilon_t. \end{aligned}$$

We are primarily interested in solutions to the above system that are stable after a particular time t. We thus directly focus on dynamic properties of deterministic part of (1)

$$\begin{aligned} y_t &= \psi_1 y_{t-1} + \psi_2 y_{t-2} + ... \\ &\quad + \sum_{i=1}^{q} \beta_i G\left(\phi_i' \left(y_{t-1}, y_{t-2}, ...\right)\right). \end{aligned} \tag{2}$$

More particularly, we are interested in finding values $y \in \mathbb{R}$ such that, after a particular time t, the solution $(y_t)_{t\in\mathbb{N}}$ to (2) satisfies $y_{t'} - y = 0$ for every $t' \geq t$. Such a value y is called a *stable value* of the system (1).

Definition. Let $(y_t)_{t\in\mathbb{N}}$ be a solution to (2). We say that the sequence $(y_t)_{t\in\mathbb{N}}$ is a *stable* solution if there exists $t_0 \in \mathbb{N}$ such that $y_{t'} - y = 0$ for every $t \geq t_0$.

3 Existence

In establishing existence of a stable solution to (1) we first analyze the case where the system receives a constant input signal, which we refer to as stable value of (1). As a simple example, it is enough for the system to start from the initial condition $y_j = y$ for every $j = 1, ..., p$ to generate the output y indefinitely, and in turn a stable solution.

Our result on existence of such equilibria of (2) will obtain under boundedness and continuity of G, together with some conditions on parameters of the AR part of the system that enable us to distinguish the general case from the orthonormal AR model, where, as shown by Proposition 2, the system (2) collapses to the linear AR model.

Assumption 1. The activation function G is continuous and bounded; i.e., there exists a constant $H > 0$ such that $|G| < H$.

Define now, for every $y \in \mathbb{N}$,

$$\Psi(y) = \psi_1 y + ... + \psi_p y + \sum_{i=1}^{q} \beta_i G\left(\phi_i'(y, ..., y)\right). \tag{3}$$

The function Ψ represents the output of the neural network after receiving as input an identical signal. It now follows that y is a stable value of (1) if and only if y is a fixed point of Ψ. We will use this property to prove existence of a stable value of (1), leading in turn to existence of a stable solution.

Existence of a stable value is proved through two modifications of Theorem 2.2 in van den Driessche and Zou [1] that distinguish between the orthonormal and non-orthonormal AR cases.

Theorem 1 *Assume that* $\sum_{i=1}^{p} \psi_i \neq 1$*, and that Assumption 1 holds. There exists a stable value of (1).*

Proof. From the above remark, existence of a stable solution is equivalent to existence of $y \in \mathbb{R}$ such that

$$y = \Psi(y).$$

Since $\sum_{i=1}^{p} \psi_i \neq 1$, existence of a stable solution is also equivalent to the existence of y such that

$$y = \frac{1}{1 - \sum_{j=1}^{p} \psi_j} \sum_{i=1}^{q} \beta_i G\left(\phi_i'(y, ..., y)\right).$$

Define the function $g : y \to \frac{1}{1-\sum_{j=1}^{p} \psi_j} \sum_{i=1}^{q} \beta_i G\left(\phi_i'(y, ..., y)\right)$. We are left to prove that g has a fixed point.

By Assumption 1, the function G is continuous and bounded, thus the function g is also continuous and bounded. Define now

$$H = \max_{y \in \mathbb{R}} |g(y)| .$$

By Brouwer's fixed point Theorem, the function g restricted to the interval $[-H, H]$ has a fixed point y^* within this interval. It follows from the above that y^* is a stable value of (1). The proof is now complete. ■

The case where $\sum_{i=1}^{p} \psi_i = 1$ is addressed next.

Proposition 2 *Assume that* $\sum_{i=1}^{p} \psi_i = 1$*. The value* y *is a stable value of* (1) *if, and only if,*

$$\sum_{i=1}^{q} \beta_i G\left(\phi_i^{'}(y, ..., y)\right) = 0.$$

Proof. Form the above remark, the value y is a stable value if, and only if,

$$y = \psi_1 y + ... + \psi_p y + \sum_{i=1}^{q} \beta_i G\left(\phi_i^{'}(y, ..., y)\right). \tag{4}$$

Since $\sum_{i=1}^{p} \psi_i = 1$, the condition (3) is equivalent to the existence of y such that

$$\sum_{i=1}^{q} \beta_i G\left(\phi_i^{'}(y, ..., y)\right) = 0.$$

The proof is now complete. ■

Proposition 2 shows that when the AR part is orthonormal, existence of a stable value is equivalent to the system (2) collapsing to a linear autoregressive model. From its proof, any y satisfying the condition $\sum_{i=1}^{q} \beta_i G\left(\phi_i^{'}(y, ..., y)\right) = 0$ is a stable value. Still, Theorem 1 does not ensure uniqueness. Consider the following trivial example in this latter case: $y_{t+1} = G(y_t)$, where $G(x) = x^3$ if $x \in [-1, 1]$, if $x < -1$ then $G(x) = -1$, and if $x > 1$then $G(x) = 1$. It is easy to see that -1, 0 and 1 are stable values for (2), and thus we have at least three stable solutions.

We next turn to providing sufficient conditions for uniqueness of a stable value of (2).

4 Uniqueness

Uniqueness is next established under the following conditions.

The first assumption is that the function G be globally Lipschitz.

Assumption 2. There exists $L > 0$ such that, for every x and x' in $\mathbb{R}$,

$$\left|G(x) - G(x')\right| \leq L\left|x - x'\right|.$$

The second assumption is on the parameters of (1).

Define first the following constants:

$$\phi \triangleq \max_{i,j}\left|\phi_{i,j}\right|,$$

and

$$\alpha \triangleq \max_{1\leq i\leq q}\left\{p\Psi_i + L\phi\sum_{j=1}^{q}\beta_j\right\}.$$

The second assumption needed for uniqueness is on the above coefficient α.

Assumption 3. $\alpha < 1$.

Now, we derive that, for every y and y' in $\mathbb{R}$,

$$\left|G\left(\phi_i^{'}(y,...,y)\right) - G\left(\phi_i^{'}(y',...,y')\right)\right| \leq L\left|\phi_i^{'}(y,...,y) - \phi_i^{'}(y',...,y')\right| \ \forall i,$$

and therefore

$$\left|\Psi(y) - \Psi(y')\right| = \left|\left|y - y'\right|\sum_{1\leq i\leq p}\Psi_i + \sum_i \beta_i\left[\begin{array}{c} G\left(\phi_i^{'}(y,...,y)\right) \\ -G\left(\phi_i^{'}(y',...,y')\right)\end{array}\right]\right|.$$

Hence, we have that

$$\begin{aligned}
\left|\Psi(y) - \Psi(y')\right| &= \left|\left|y - y'\right|\sum_{1\leq i\leq p}\Psi_i + \sum_i \beta_i\left[\begin{array}{c} G\left(\phi_i^{'}(y,...,y)\right) \\ -G\left(\phi_i^{'}(y',...,y)\right)\end{array}\right]\right| \\
&\leq \left|\left|y - y'\right| p\max_{1\leq i\leq q}\Psi_i + \sum_i \beta_i L\left|\begin{array}{c} \phi_i^{'}(y,...,y) \\ -\phi_i^{'}(y',...,y')\end{array}\right|\right| \\
&\leq \max_{1\leq i\leq q}\left\{p\Psi_i + L\phi\sum_{j=1}^{q}\beta_j\right\}\left|y' - y\right|
\end{aligned}$$

Thus, we have established that

$$|\Psi(y) - \Psi(y')| \leq \alpha |y - y'|$$

showing that $\Psi(\cdot)$ is a contraction operator since, by Assumption 3, $\alpha < 1$.

We can now state the next result of the paper.

Theorem 3 *Under Assumptions 2 and 3, there exists a unique stable value of* (2)*.*

Proof. The uniqueness of a stable value is a straightforward consequence of the above analysis and of the Contraction Mapping Theorem as in Luenberger [3], or in Kantorovich and Akilov [2]. ■

We next study the global stability of the above value.

5 Global stability

We now turn to studying global stability of the unique stable solution found in Theorem 3. under Assumptions 2-3.

We first say that a value y is *globally stable* if, for every possible vector of initial conditions, every solution to (2) converges to y.

Theorem 4 *Under Assumptions 2 and 3, there exists a unique globally stable value. Moreover, the globally stable value is the stable value.*

Proof. Let y be the unique stable value found in Theorem 3, and let $(y_t)_{t\in\mathbb{N}}$ be the solution to (2) for any arbitrary vector of initial values $(y_1, ..., y_p)$.

For any $t \geq p$, we have that

$$\begin{aligned} |y - y_t| &= |h(y, ..., y) - h(y_{t-1}, ..., y_{t-p})| \\ &\leq \alpha. \max_{1\leq i\leq q} |y - y_{t-i}|, \end{aligned} \tag{5}$$

by a reasoning similar to the proof of Theorem 3.

Let $i_1 \in \arg\max_{1\leq i\leq q} |y - y_{t-i}|$. In turn, we also have that

$$\begin{aligned} |y - y_{t-i_1}| &= |h(y, ..., y) - h(y_{t-i_1-1}, ..., y_{t-i_1-p})| \\ &\leq \alpha. \max_{1\leq i\leq q} |y - y_{t-i_1-i}|. \end{aligned} \tag{6}$$

Plugging (6) into (5) leads to

$$|y - y_t| \le \alpha^2 . \max_{1 \le i \le q} |y - y_{t-i_1-i}| .$$

We can iterate the above process to get:

$$|y - y_t| \le \alpha^{[\frac{t}{p}]} . \max_{1 \le i \le q} |y - y_i| , \tag{7}$$

where $[\frac{t}{p}]$ is the greatest of all integers n such that $p.n \le t$.

Since $\alpha < 1$ by Assumption 3, it follows that the expression $\alpha^{[\frac{t}{p}]}$ converges to 0 as t converges to infinity. Thus the right-hand side of (7) converges to 0 as t converges to infinity, and from (7) we finally get that

$$|y - y_t| \overset{t \to \infty}{\longrightarrow} 0.$$

The proof is now complete. $\blacksquare$

From inequality (7) in the above proof, we can see that convergence toward the stable occurs at exponential rate. In other words, convergence of the AR-NN is fast, and thus learning takes place quickly.

6 Conclusions

This paper shows that the boundedness and continuity of the activation function of the hidden units in the multi layer perceptron (MLP) is a sufficient condition for the existence of a stable solution of AR-NN processes. Global Lipschitzness of the MLP together with some restrictions on the parameters of the system ensures uniqueness of the stable solution; in this case, the system is proven to globally converge towards the unique stable value at exponential rate.

References

[1] van den Driessche, P., and X. Zou (1998) "Global Attractivity in Delayed Hopfield Neural Network Models," *SIAM Journal on Applied Mathematics,* 58 (6), 1878-1890.

[2] Kantorovich, L., and G. Akilov (1982) *Functional Analysis*, Oxford: Pergamon Press.

[3] Luenberger, D. (1969) *Optimization by Vector Space Methods*, New-York: Wiley.

[4] Trapletti, A., F. Leisch and K. Hornik (2000) "Stationary and Integrated Autoregressive Neural Network Processes," *Neural Computation,* 12 (10), 2427-2450.

[5] Beale, R., and T. Jackson (1990) *Neural Computing, an Introduction*, Adam Hilger, IOP Publishing Ltd: Bristol.

[6] Kohonen, T. (1988) "An introduction to neural computing," *Neural Networks*, 1 (1), 3-16.

[7] Poff, L.N., S. Tokar, and P. Johnson (1996) "Stream hydrological and ecological responses to climate change assessed with an artificial neural network," *Limnology and Oceanography*, 41 (5), 857.

[8] Zirilli, J.S. (1996) *Financial Prediction Using Neural Networks*, International Thomson Publishing,